REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 31,2002	Final Report: 16 Mar 1998 - 31 Dec 2001

4. TITLE AND SUBTITLE
Developing and Understanding Methods for Large-Scale Nonlinear Optimization

5. FUNDING NUMBERS
DAAG55-98-1-0176

6. AUTHOR(S)
Principal Investigator: Robert B. Schnabel
Co-Principal Investigator: Richard H. Byrd

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Colorado
Boulder, Colorado 80305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
37302.15-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
The overall area of this research is the development large-scale optimization methods. The primary research objective is to develop efficient and effective large-scale global optimization methods for determining the structure of polymers and proteins. An accompanying objective development of efficient parallel large-scale global optimization methods.
We have made major advances in the applicability of our methodology. The culmination of these advances was our participation over the summer in the fourth CASP (Critical Assessment of Techniques for Protein Structure Prediction) competition, where we have attempted blind prediction of eight proteins of arbitrary complexity including mixed alpha helices and beta sheets.
Our results were excellent, standing in the top quartile of all groups for the 3 proteins that were in the top 15% of difficulty, and the best result of all groups for the hardest protein we attempted with 242 amino-acids.
Enabling us to reach this stage have been advances in several areas including the ability to handle beta sheets, and the development of new biasing techniques.

14. SUBJECT TERMS
numerical optimization, global optimization, protein folding

15. NUMBER OF PAGES
6

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (FROM - TO)
31-05-2002		16-03-1998 to 31-12-2001

4. TITLE AND SUBTITLE
Developing and Understanding Methods for Large Scale Nonlinear Optimization
Unclassified

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME AND ADDRESS
University of Colorado
Boulder
Boulder, CO80305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS
,

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APUBLIC RELEASE
,

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Our main activity in this grant has been the development and testing of techniques for solving global optimization problems for determining the structure of proteins and polymers. The problem is to find the lowest energy configuration of a protein or other polymer. This problem is a global optimization problem because it has a huge number of local minimizers. In addition, locating the lowest (global) minimizer is very difficult. For proteins, the solution of this problem would represent a solution to the wellknown protein-folding problem.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19. NAME OF RESPONSIBLE PERSON
			Public Release	7	Rike, Jack jrike@dtic.mil
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number DSN

FINAL REPORT for

Army Research Office Grant DAAG55-98-1-0176
Developing and Understanding Methods for
Large-Scale Nonlinear Optimization

16 March 1998 - 31 December 2001

Principal Investigator: Robert B. Schnabel
Co-Principal Investigator: Richard H. Byrd

1. Statement of the Problem Studied

This research is focussed on methods for the numerical solution of nonlinear optimization problems. Optimization problems constitute an important class of mathematical problems whose computer solution is desired by scientists and engineers. The need to solve such problems arises in a huge variety of computer simulations and data analyses, with examples including optimal design of aircraft and spacecraft for low cost or high efficiency, optimal control of chemical processes, and calculation of the native states of biological molecules. In recent years, as computer power has increased, the feasibility of investigating optimal designs has increased, and along with this the need for improved optimization methods has grown. This requires fundamental research in optimization algorithms, especially for classes of large-scale problems. In addition, much of this computation must utilize the fastest available computers, which increasingly are parallel computers. Thus the development of optimization methods that are well-suited to parallel computers is also a pressing need. The research funded by the grant was in 3 areas: development of global optimization methods for prediction of protein structure by potential energy minimization, the development of interior point methods for large-scale constrained optimization, and the development of tensor methods for large-scale systems of nonlinear equations.

One of the main emphases of this research program in recent years has been the development of global optimization methods for the solution of large molecular configuration problems. Global optimization means finding the lowest minimizer of a nonlinear function that may have numerous local minimizers. The objective of this research is to develop methods that are reliably capable of solving difficult, large nonlinear global optimization problems. Our current work is investigating methods to find the native configurations of proteins and polymers. This is a problem of great importance in science; it includes the well-known protein folding problem as well as the investigation of polymers that are used to make new materials. The resultant optimization problems are very difficult because they have extremely large numbers of local minimizers that have similar function values.

Another main focus of our research is the development of new algorithms for large-scale unconstrained and constrained optimization problems, including limited-memory methods for problems with many thousands of variables, and interior point methods for nonlinearly constrained problems. We are also investigating theoretical convergence issues for optimization algorithms that have practical consequences in how problems are formulated, and new approaches to one of the key numerical linear algebra subproblems that is particular to optimization algorithms.

2. Summary of the Most Important Results

Our main activity in this grant has been the development and testing of techniques for solving global optimization problems for determining the structure of proteins and polymers. The problem is to find the lowest energy configuration of a protein or other polymer. This problem is a global optimization problem because it has a huge number of local minimizers. In addition, locating the lowest (global) minimizer is very difficult. For proteins, the solution of this problem would represent a solution to the well-known protein-folding problem.

In previous research periods, we have developed a stochastic/perturbation approach for solving global optimization problems from molecular chemistry. There are four keys to this approach. The first is a large scale global optimization methodology that performs small-scale global optimizations with only a small number of parameters variable and the remaining parameters temporarily fixed, followed by local minimizations with all parameters varying, at each stage of the global optimization procedure. The second is the incorporation of a new, efficient approach towards smoothing the objective function in the global optimization framework. Initially these have been the backbone of the approach. More recently two other aspects have become key to doing work on realistic protein targets. One of these is the incorporation of predictions from secondary structure prediction methods in the initial phase of our algorithm, to produce starting configurations with reasonably good secondary structure. The final one is work with our chemistry partners to use the mismatch between simulation and experimental evidence to continue to refine the mathematical energy model upon which our global optimization approach relies.

Originally, our research had been applied primarily to molecular clusters, and to small, alpha-helical proteins. We had developed good biasing methods for creating predicted alpha-helical secondary structure at the start of our method, and, in the last research period, had shown that our global optimization approach could do a reasonably good job of predicting the full tertiary structure of several helical proteins of about 70 amino acids. We had also demonstrated the effectiveness of our smoothing approach.

In the course of this grant, we evolved our approach to be able to handle proteins with arbitrary structure. The crux of this issue, for us and other groups doing related research, is the ability to handle beta-sheets. Beta-sheets are the other main type of secondary structure in proteins. However they are far less local than alpha-helices. While alpha-helices are continuous, beta-sheets are formed by contiguous strands that can be arbitrarily far apart. Secondary structure prediction programs can predict the strands with good accuracy, but they do not predict which strands are bonded together, nor the parallel or anti-parallel orientation of those bonds.

One main component of our research was the development of a biasing function that, given predictions of which amino-acids are bonded together to form the beta-sheet, influences the protein to form these bonds. Biasing functions are simply penalty functions from optimization that are added to the energy function. We constructed a function that is a combination of a sigmoid at low distances and a linear function at higher distances, to balance the need to form the bonds but not to overly bias. Along with this, we built upon existing software from the bio-chemistry community to construct techniques to predict the several most likely combinations of the predicted beta strands into beta sheets. The tests we describe later in this section, as well as earlier, simpler tests, clearly demonstrated the viability of these new approaches.

Simultaneously with the beta sheet research, we continued to refine the heart of the global optimization algorithm. As the protein sizes increase, the selection of roughly 5-6 dihedral angles (out of

200-500) at each stage to be the parameters in the small scale global optimization becomes even more crucial. And, not only should these be the parameters whose variation can lead to improvements in the tertiary structure, but they also must be a set that can be varied without destroying good secondary structure. We have developed new approaches to selecting these parameters that emphasize selecting from portions of the protein that are not parts of the regions of secondary structure. We also have begun to develop ways to selecting a set of angles from turns connecting beta sheets so as not to destroy the beta sheet.

Equally important is the ability to determine structural different proteins to work upon. This issue generalizes to any global optimization problem where we want to spread our effort over the search space. During this grant we formulated and implemented a clustering technique that takes all the currently active configurations and groups them into clusters of similarly shaped configurations. Then our algorithm only proceeds with one configuration from any given cluster at once. The clustering also is used to determine when to stop our algorithms, based upon whether the number of clusters still is growing, or not.

An important milestone in our research was our participation in the fourth Critical Assessment of Techniques for Protein Structure Prediction (CASP4) competition in summer 2000. This competition, held every two years in the bio-chemistry community, invites any interested groups to blindly predict the structure of proteins that are about to be experimentally analyzed. About 170 groups entered this year. These groups utilize many different approaches, most based upon comparison to the structures of known proteins. Our approach, which only uses secondary structure prediction but not sequence matching, is at the pure end of the spectrum and is particularly important for predicting "new folds" that do not closely match known proteins. We spent the summer predicting eight proteins, with sizes ranging from 56 to 242 amino acids. Several of these were helical but the majority were mixed alpha-beta proteins. Overall our group did quite well in this competition. Our predictions were in the top quartile of all groups for the 3 proteins that were in the top 15% of difficulty, and the best result of all groups for the hardest protein we attempted with 242 amino-acids.)

Based in part on this the results of this competition we have been able to make several improvements in the algorithm, particular in the area of predicting beta sheets.

In addition, we have continued our collaborations with Jorge Nocedal at Northwestern University on the development of a robust algorithm for nonlinearly constrained optimization. During this research period we have continued to develop a software package for this problem, KNITRO, and conducted extensive comparative tests of it and several other leading packages. The tests show that no one method is preferable in all situations, but that KNITRO is a very competitive method. We have added several enhancements to KNITRO including a strictly feasible version, and a quasi-Newton version.

During this research period we also began work on new tensor methods for very large scale of nonlinear equations. We have developed a new approach for iteratively solving the tensor model that avoids the cost of a second backsolve that the previous approach had. We have also developed a new curvilinear line search for tensor methods that eliminates the need to use the tensor and Newton direction separately and which produces monotonic descent on the tensor model.

3. List of All Publications and Technical Reports

Submitted but not yet accepted:

R.H. Byrd and H. Khalfan, "Analysis of a symmetric rank-one trust region method for constrained minimization", submitted for journal publication.

H. Khalfan, R.H. Byrd, and R. Schnabel, "Retaining convergence properties of trust region methods without extra gradient evaluations" submitted for journal publication.

A. Azmi, R. Byrd, E. Eskow and R. Schnabel, "New smoothing techniques for global optimization in solving for protein conformation", submitted for publication.

R. Byrd, J. Nocedal and R. Waltz, "Feasible Interior Methods Using Slacks for Nonlinear Optimization," submitted for journal publication.

R. Byrd, M. Marazzi and J. Nocedal, "On the Convergence of Newton Iterations to Non-Stationary˜Points", submitted for journal publication.

Accepted but not yet published:

D. Feng and R. Schnabel, "Local convergence analysis of tensor and SQP methods for singular constrained optimization", to appear in *SIAM Journal on Optimization.*

Published in peer-reviewed journals

A. Bouaricha and R. Schnabel, "Tensor methods for large sparse systems of nonlinear equations", *Mathematical Programming* 82, 1998, pp. 377-400.

A. Bouaricha and R. Schnabel, "Tensor methods for large sparse nonlinear least squares problems", *SIAM Journal on Scientific Computing.* 21, 1999, pp. 1199-1221.

R.H. Byrd, J.C. Gilbert and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Mathematical Programming* 89, 2000, pp. 149-185.

R.H. Byrd, M. E. Hribar and J. Nocedal, "An Interior Point Algorithm for Large Scale Nonlinear Programming," *SIAM Journal on Optimization* 9, 1999, pp. 877-900.

R. Byrd and J. Nocedal, "Active set and interior point methods for nonlinear optimization", *Documenta Mathematica, Vol III, Journal der Deutschen Mathematiker-Vereinigung,* 1998.

S. Crivelli, B. Bader, R. Byrd, E. Eskow, V. Lamberti, R.Schnabel and T. Head-Gordon "A physical approach to protein structure prediction", *Biophysical Journal* 82, 2002, pp. 36-49

S. Crivelli, R.H. Byrd, E. Eskow, R. Schnabel, R. Yu, T. Phillips and T. Head-Gordon, "A global optimization strategy for predicting alpha-helical protein tertiary structure", *Computers and Chemistry 24* (May 2000), pp. 489-497.

R. Schnabel and E. Eskow, "A revised modified Cholesky factorization" *SIAM Journal on Optimization,* Sept. 1999

C. Shao, R. Byrd, E. Eskow and R. Schnabel, "Global optimization for molecular clusters using a new smoothing approach", *Journal of Global Optimization* 16, 2000, pp. 167-196.

Y. Xie and R. Byrd, "Practical update criteria for reduced Hessian successive quadratic programming algorithms," *SIAM Journal on Optimization.* 9, 1999, pp. 578-604.

Published in conference proceedings

A. Azmi, R. Byrd, E. Eskow, R. Schnabel, S. Crivelli, T. Philip and T. Head-Gordon, "Predicting Protein Tertiary Structure Using a Global Optimization Algorithm with Smoothing", *Optimization in Computational Chemistry and Molecular Biology: Nonconvex Optimization and Its Applications,* C.A. Floudas and P.M. Pardalos, eds., Kluwer Academic Publishers, pp. 1-18, 2000.

S. Crivelli, T. Head-Gordon, R. Byrd, E. Eskow and R. Schnabel, "A Hierarchical Approach for Parallelization of a Global Optimization Strategy for Protein Structure Prediction", *Proceedings of Euro-Par '99,* Lecture Notes in Computer Science Series, Springer-Verlag.

C. Shao and R. Schnabel, "A Task Migration System for Parallel Scientific Computations in Heterogeneous NOW Environments", *"Proceedings of the Ninth SIAM Conference on Parallel Processing for Scientific Computing",* B. Hendrickson, K. Yelick, C. Bischof, I. Duff, A. Edelman, G. Geist, M. Heath, M. Heroux, C. Koelbel, R. Schrieber, R. Sincovec, and M. Wheeler, eds., CD-ROM, ISBN 8-89871-435-4.

Presented at meetings but not published

R. Byrd, "Behavior of an interior point method for general nonlinear optimization", invited minisymposium talk, SIAM National Meeting, Toronto, July 1998.

R. Byrd, "Using global optimization to find molecular structures", Mathematics Department Colloquium, Instituto Tecnologico Autonomo de Mexico, Mexico City, February, 8 1999.

R. Schnabel, "Predicting protein tertiary structure using a global optimization algorithm with smoothing", SIAM Conference on Optimization Atlanta, May 1999.

R. Byrd, "Use of trust regions in interior point methods for nonlinear optimization", SIAM Conference on Optimization Atlanta, May 1999.

R. Byrd, "Step computation in a trust region interior point method," First Workshop on Nonlinear Optimization, University of Coimbra, Portugal, October 18, 1999.

R. Byrd, "False Convergence in Optimization Algorithms", International Conference on Mathematical Programming, Atlanta, August 2000.

E. Eskow, "A Stochastic_Perturbation Global Optimization Approach to Protein Structure Prediction", International Conference on Mathematical Programming, Atlanta, August 2000.

R. Schnabel, "Optimization applied to Protein Folding: Interrelationships and Recent Progress", International Conference on Mathematical Programming, Atlanta, August 2000.

R. Schnabel, "Protein Structure Prediction by Global Optimization Utilizing Secondary Structure Prediction", SIAM Conference on Scientific Computing, Washington D.C., Sept. 2000.

Richard Byrd, "Reliability of Nonlinear Optimization Algorithms", minisymposium talk presented at the INFORMS International Meeting, June 2001.

Richard Byrd, "Causes of Failure in Optimization Algorithms", Seventh US-Mexico Workshop in Numerical Analysis, Merida, Mexico, January 2001.

4. Participating Scientific Personnel

Betty Eskow, Professional Research Assistant

Aqil Azmi, Ph.D 1998

Anna Szczyrba. M.S. 1998

Vincent Lamberti, M.S. 2000

Brett Bader, doctoral student

Elizabeth White, doctoral student